

# **MINERvA** Analysis and Statistical Methods

Laura Fields



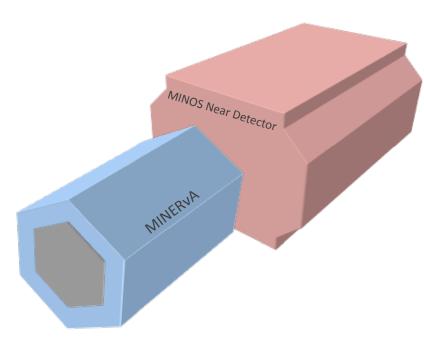
#### **Outline**

- MINERVA
  - What
  - How
- Statistical problems and how we've solved them:
  - Propagating systematic uncertainties
  - Constraining the NuMI flux
  - Unfolding





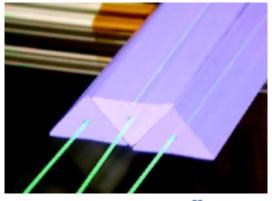
A neutrino detector in the NuMI beam at Fermilab designed to study v<sub>μ</sub> and v<sub>e</sub> interactions with nuclei (and to compare across different nuclei)



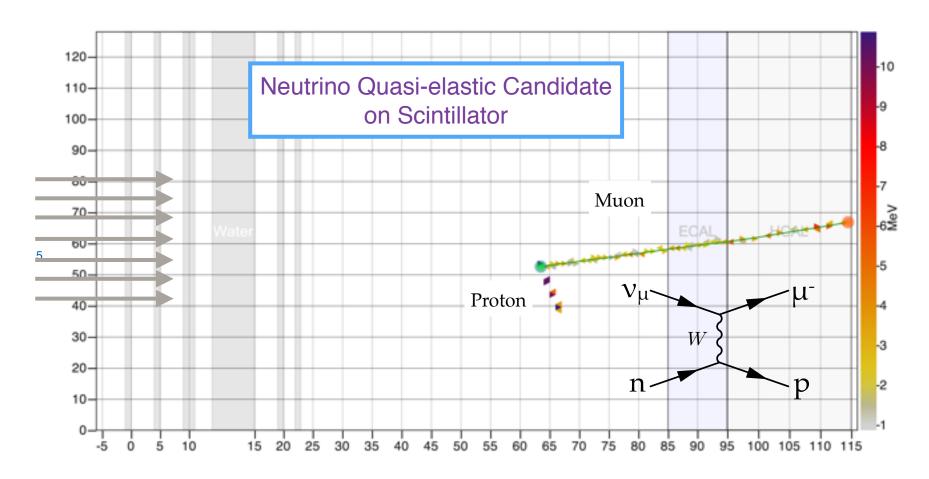




- Made of > 30,000 strips of plastic scintillator interspersed with other materials
- Scintillator creates light when charged particles move through it



# A sample neutrino interaction in MINERvA:





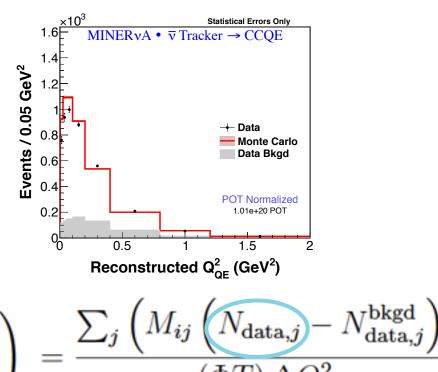
Laura Fields I MINERvA 19/09/16

- MINERvA measures cross sections probabilities that some processes will happen
- Often we measure differential cross sections with respect to some variable
- For example, for our first measurement, we measured quasi-elastics (the example on the previous slide) as a function of a variable called Q<sup>2</sup>.
- Here is the basic recipe for a cross section:

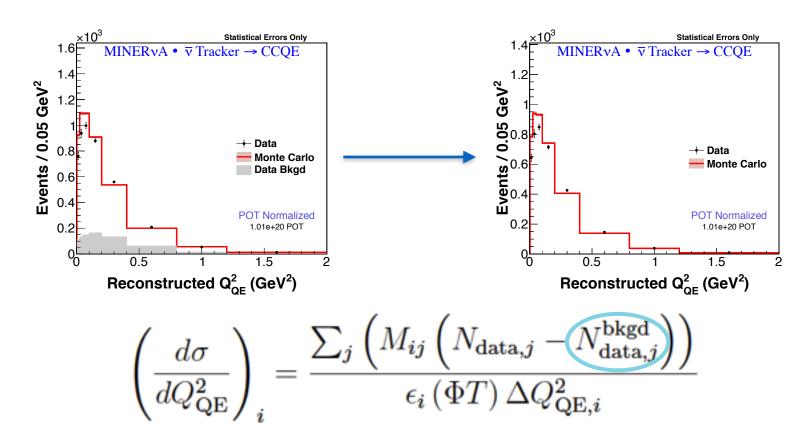
$$\left(\frac{d\sigma}{dQ_{\text{QE}}^{2}}\right)_{i} = \frac{\sum_{j} \left(M_{ij} \left(N_{\text{data},j} - N_{\text{data},j}^{\text{bkgd}}\right)\right)}{\epsilon_{i} \left(\Phi T\right) \Delta Q_{\text{QE},i}^{2}}$$



 We start by a selecting a sample of events enriched with whatever process we want to measure and bin them in the variable we want to measure

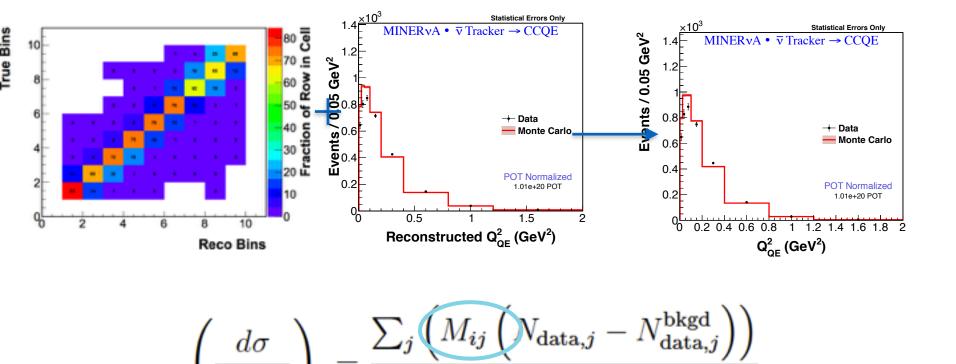


 We then subtract our best estimate of backgrounds (almost always constrained with a data fit or sideband)



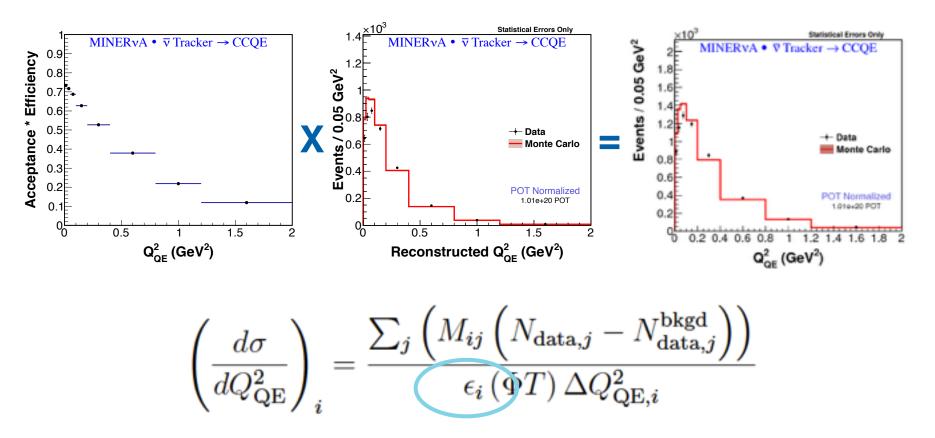


 We unfold to correct for detector smearing in the variable we are measuring (more on this shortly):



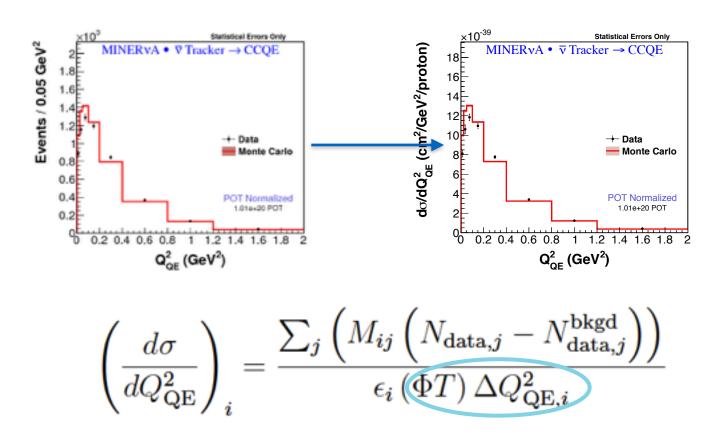


 We then correct for analysis efficiency (events lost due to our analysis cuts) and detector acceptance (events lost due to detector geometry):

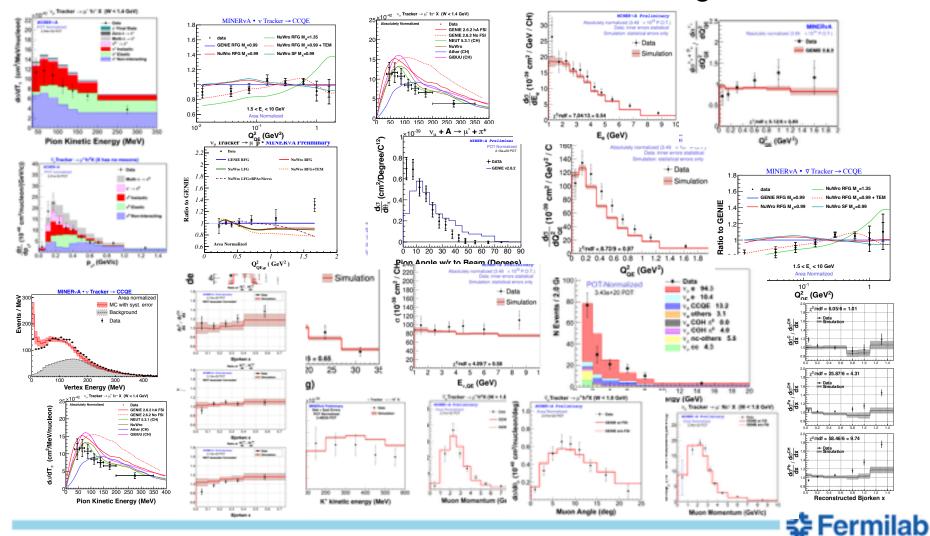




• Finally, we divide by the number of targets (usually nucleons) in our detector and the number of neutrinos in our beam (more on this soon):



And then we do that, over and over again

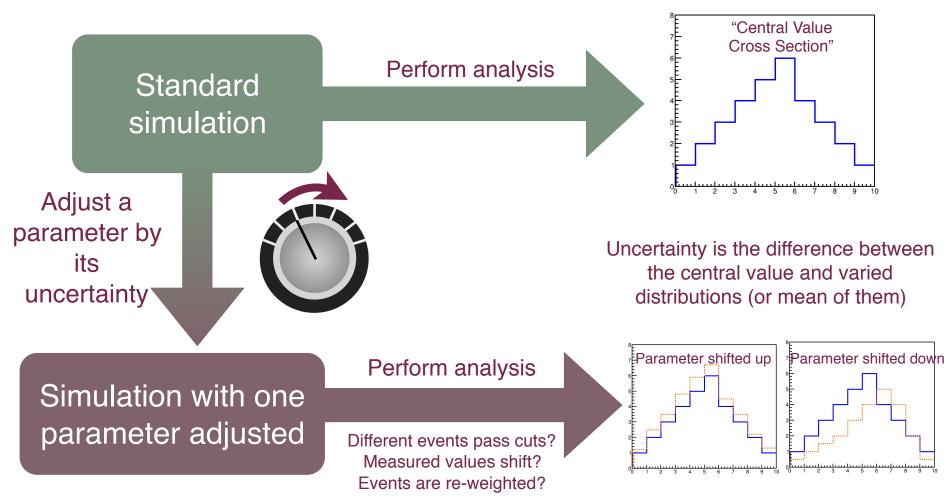


12 Laura Fields I MINERvA 19/09/16

- As everyone in this room definitely knows, a cross section measurement is meaningless without an error bar (and correlation matrix!)
- MINERvA cross section measurements have systematic uncertainties from many sources, e.g.
  - Neutrino flux
  - Mass of detector
  - Energy scales of reconstructed particles/energy
  - Tracking efficiencies
  - Models of neutrino interactions and FSI
  - Deadtime

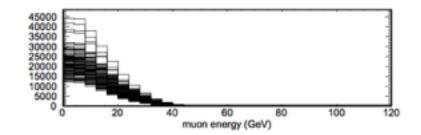


We assess systematic uncertainties in the usual way:

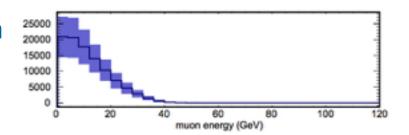


- When assessing systematic uncertainties due to many correlated parameters, we use the "many-universe technique"
  - Uncertain parameters are selected randomly from their probability distributions
  - This is done many times (100-1000)

For each set of parameters (ie in each "universe"), a new simulated distribution is produced corresponding to that universe



The RMS of of the predicted distribution in a particular bin becomes the uncertainty on that bin



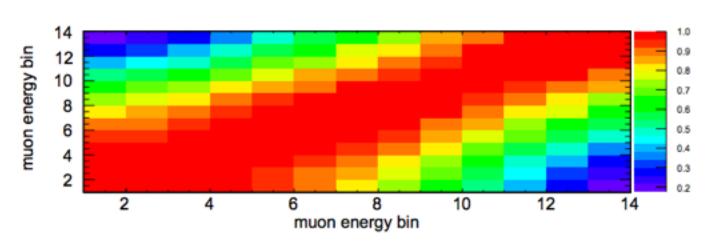


Correlations are also extracted from the universes

cross section in bin k of ith universe

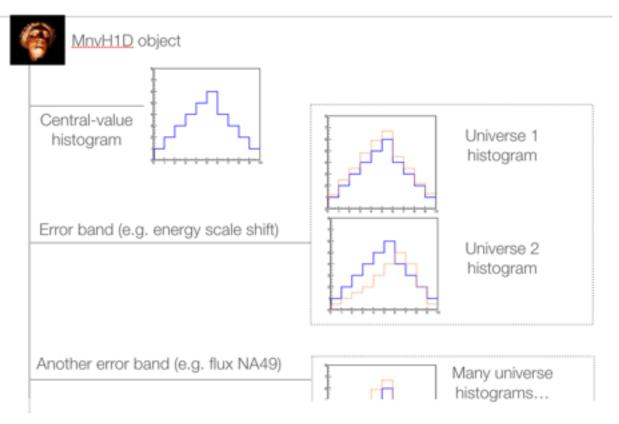
$$\operatorname{cov}(j,k) = \frac{1}{N} \sum_{i}^{\operatorname{histos}} (\nu_j - n_{ij}) (\nu_k - n_{ik})$$

cross section in bin j of central value



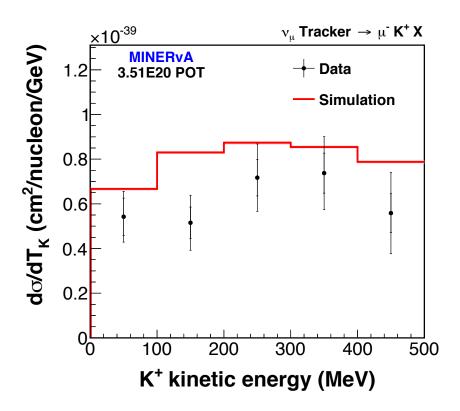


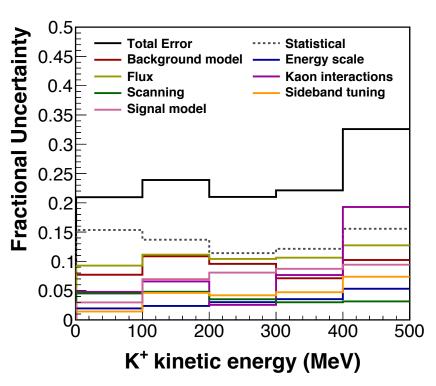
 The multi-universe method is so critical to MINERvA's analysis methods that we have created extensions of ROOT histogram object to facilitate it:





 The MnvH1D (and MnvH2D) class makes computing and plotting systematic uncertainties straightforward

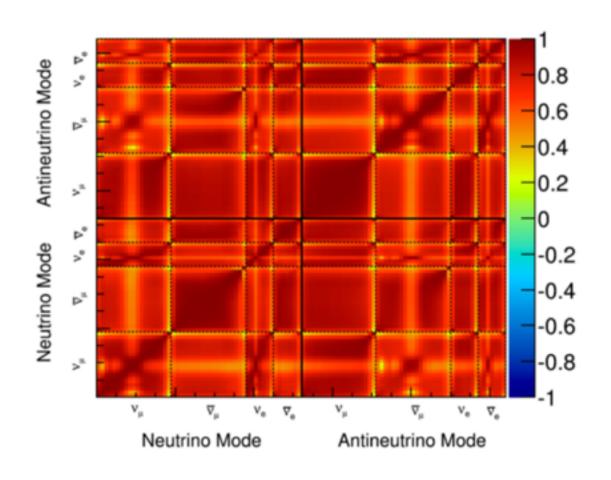




Phys. Rev. D 94, 012002 (2016)



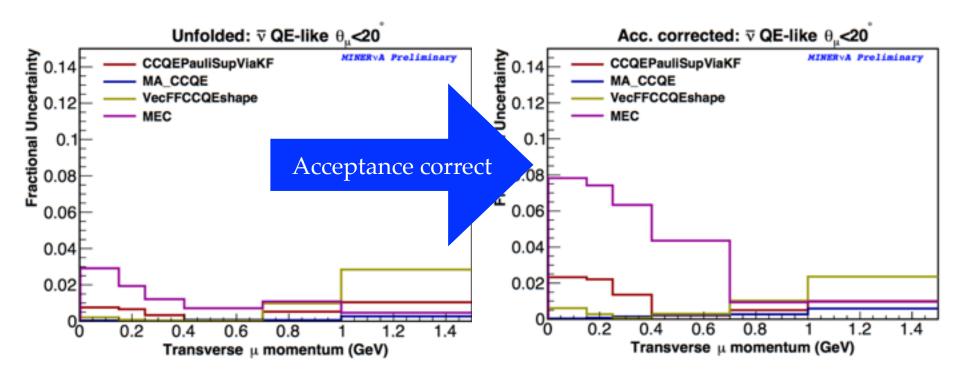
As well as correlations:



Covariance matrix of flux in neutrino energy bins requested by DUNE that was easy since we had flux MnvH1D's available



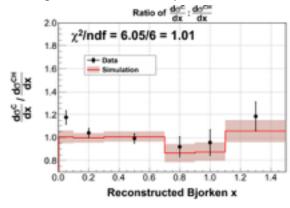
 MnvH1D construction starts at the beginning of the analysis, and is propagated through background subtraction, efficiency correction, etc.
 This is extremely useful for understanding where systematics become a problem:

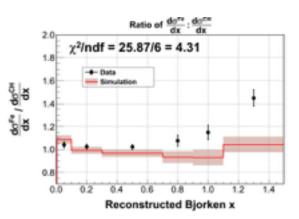


C. Patrick FNAL W&C Seminar, 17 June 2016

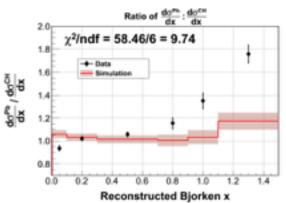


 It also makes taking ratios of measurements with correlated uncertainties extremely straightforward (ratios are calculated universeby-universe):





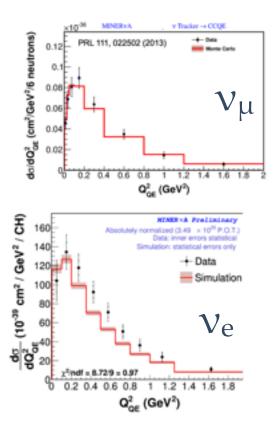
charged-current inclusive cross section ratios of different nuclei to scintillator

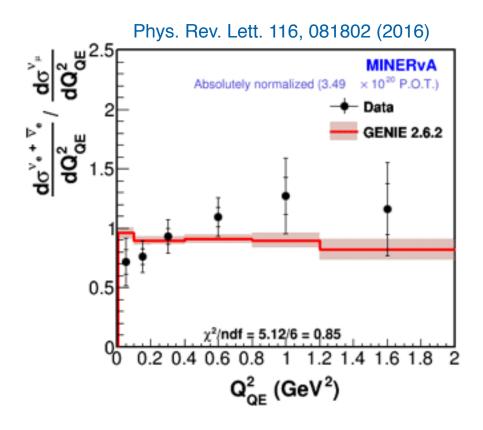


Phys. Rev. Lett. 112, 231801 (2014)



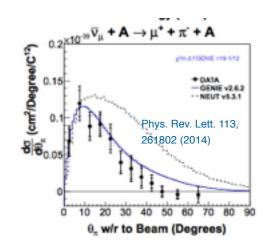
 It also makes taking ratios of measurements with correlated uncertainties extremely straightforward (ratios are calculated universeby-universe):







 We have typically provided our results to the world in the form of cross section values, errors and correlations matrix:



But given their utility, we are considering eventually making the MnvH1D's public

Degrees	0 - 5	5 - 10	10 -15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 60	60 - 70
0 - 5	0.0552	0.0801	0.0547	0.0661	0.0640	0.0369	0.0322	0.0227	0.0184	0.0014	-0.0006	0.0042
5 - 10	0.0801	0.1166	0.0797	0.0961	0.0932	0.0540	0.0470	0.0333	0.0269	0.0022	-0.0008	0.0061
10 -15	0.0547	0.0797	0.0547	0.0659	0.0641	0.0370	0.0323	0.0229	0.0187	0.0016	-0.0004	0.0043
15 - 20	0.0661	0.0961	0.0659	0.0796	0.0774	0.0446	0.0390	0.0277	0.0226	0.0019	-0.0005	0.0052
20 - 25	0.0640	0.0932	0.0641	0.0774	0.0760	0.0438	0.0385	0.0276	0.0228	0.0022	-0.0003	0.0052
25 - 30	0.0369	0.0540	0.0370	0.0446	0.0438	0.0255	0.0223	0.0160	0.0132	0.0013	-0.0001	0.0030
30 - 35	0.0322	0.0470	0.0323	0.0390	0.0385	0.0223	0.0196	0.0141	0.0117	0.0012	-0.0001	0.0027
35 - 40	0.0227	0.0333	0.0229	0.0277	0.0276	0.0160	0.0141	0.0102	0.0086	0.0010	0.0000	0.0019
40 - 45	0.0184	0.0269	0.0187	0.0226	0.0228	0.0132	0.0117	0.0086	0.0074	0.0009	0.0002	0.0017
45 - 50	0.0014	0.0022	0.0016	0.0019	0.0022	0.0013	0.0012	0.0010	0.0009	0.0002	0.0001	0.0002
50 - 60	-0.0006	-0.0008	-0.0004	-0.0005	-0.0003	-0.0001	-0.0001	0.0000	0.0002	0.0001	0.0001	0.0001
60 - 70	0.0042	0.0061	0.0043	0.0052	0.0052	0.0030	0.0027	0.0019	0.0017	0.0002	0.0001	0.0004

TABLE XVII: Anti-neutrino  $d\sigma/d\theta_{\pi}$  flux systematic covariance matrix  $\times 10^{-81}$ 

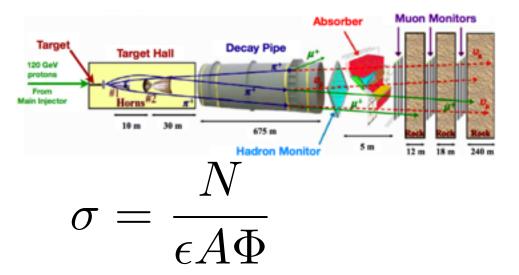
Degrees	0 - 5	5 - 10	10 -15	15 - 20	20 - 25	25 - 30	30 - 35	35 - 40	40 - 45	45 - 50	50 - 60	60 - 70
0 - 5	0.0931	0.0923	0.0859	0.0256	0.0682	0.0426	0.0683	0.0556	0.0543	0.0173	0.0259	0.0839
5 - 10	0.0923	0.1068	0.0968	0.0311	0.0804	0.0525	0.0815	0.0675	0.0671	0.0246	0.0319	0.0988
10 -15	0.0859	0.0968	0.1070	0.0336	0.0804	0.0567	0.0837	0.0692	0.0696	0.0300	0.0353	0.0917
15 - 20	0.0256	0.0311	0.0336	0.0280	0.0347	0.0295	0.0354	0.0330	0.0334	0.0220	0.0182	0.0305
20 - 25	0.0682	0.0804	0.0804	0.0347	0.0848	0.0550	0.0756	0.0663	0.0659	0.0331	0.0341	0.0807
25 - 30	0.0426	0.0025	0.0567	0.0295	0.0550	0.0045	0.0566	0.0514	0.0519	0.0313	0.0280	0.0524
30 - 35	0.0583	0.0815	0.0837	0.0354	0.0756	0.0566	0.0856	0.0678	0.0682	0.0349	0.0352	0.0803
35 - 40	0.0556	0.0675	0.0692	0.0330	0.0663	0.0514	0.0678	0.0690	0.0612	0.0339	0.0321	0.0674
40 - 45	0.0543	0.0671	0.0696	0.0334	0.0659	0.0519	0.0682	0.0612	0.0693	0.0349	0.0326	0.0664
45 - 50	0.0173	0.0246	0.0300	0.0220	0.0331	0.0313	0:0349	0.0339	0.0349	0.0310	0.0197	0.0248
50 - 60	0.0259	0.0319	0.0353	0.0182	0.0341	0.0280	0.0352	0.0321	0.0326	0.0197	0.0216	0.0321
60 - 70	0.0859	0.0988	0.0917	0.0305	0.0807	0.0524	0.0803	0.0674	0.0664	0.0248	0.0321	0.1003

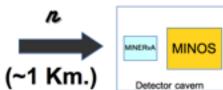
TABLE XVIII: Anti-neutrino  $d\sigma/d\theta_{\pi}$  non-flux systematic covariance matrix  $\times 10^{-81}$ 



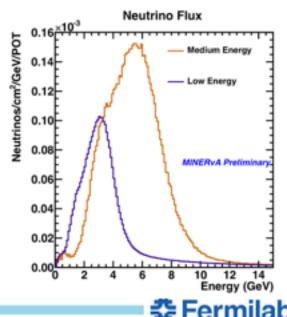


One of our most important systematic uncertainties: the neutrino flux

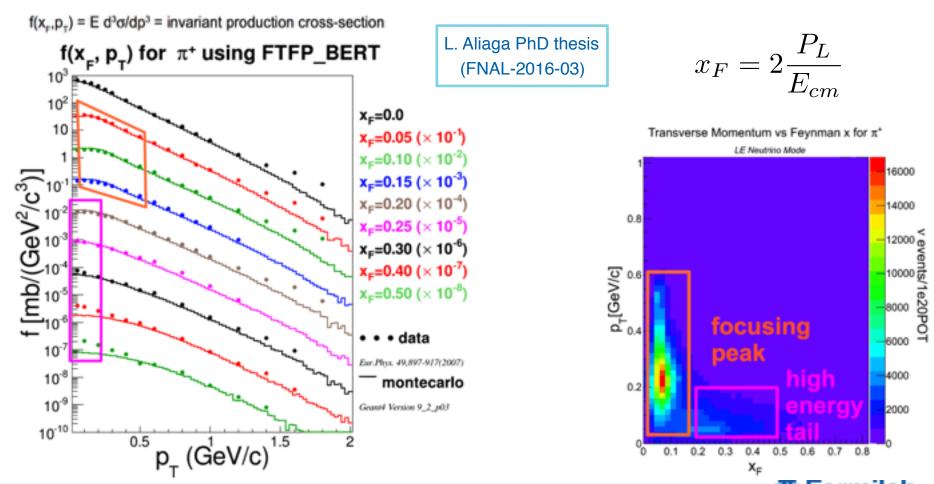




- Flux integral is in the denominator of all of our cross section; also is the starting point of simulations used to estimated backgrounds, acceptance and smearing
- Flux simulation starts with a GEANT4 simulation of the NuMI beam line (G4NuMI)



One problem: Geant4 does not always agree with external data:



- We force the simulation to match external data
- How this works in practice:
  - Complete information about cascades leading to a neutrino is recorded for each proton on target and stored in the flux tuple
  - Including interaction materials and ancestor kinematics
  - In MINERvA analyses, neutrino events are weighted by:

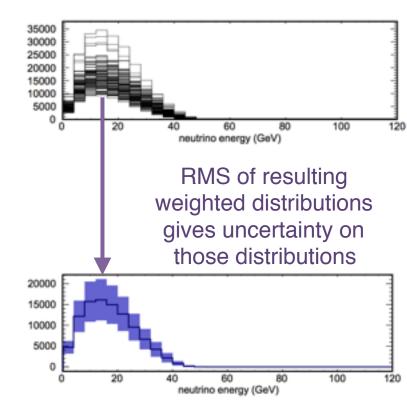
$$w_{\mathrm{HP}} = rac{f_{\mathrm{Data}}(x_F, p_T, E)}{f_{\mathrm{MC}}(x_f, p_T, E)} \qquad f = E rac{d^3 \sigma}{dp^3}$$



Weights for events with multiple interactions are the product of individual interaction weights

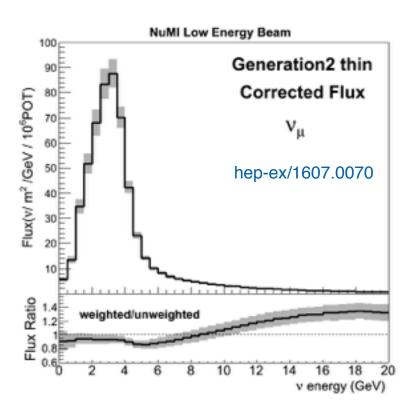


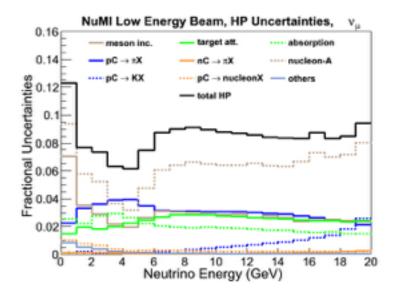
- Uncertainties on the external data constraints are propagated to uncertainties on our flux using the many universes method:
- For each event, in addition to the central value weights, we also store many (~1000) weights constructed from data cross sections varied according to their uncertainties (taking into account correlations)





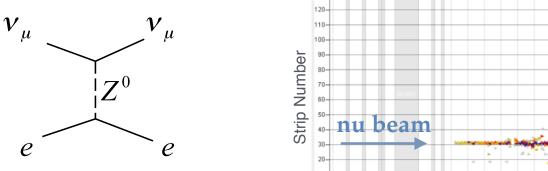
• The resulting flux / uncertainties:



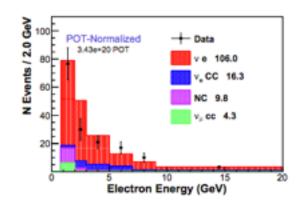


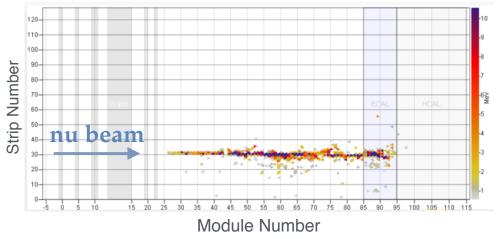


• Also pioneering use of a "new" standard candle for flux estimation: neutrino scattering on electrons:  $\sigma = \frac{1}{2}$ 



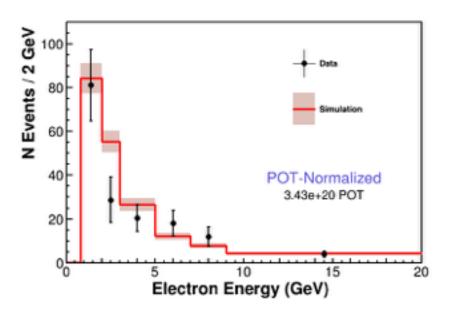
Phys. Rev. D 93, 112007 (2016)





- Well understood electroweak process
- Signal in MINERvA is a single electron moving in the beam direction
- Process cross section is smaller than nucleus scattering by a factor of 2000 -> statistically limited





Phys. Rev. D 93, 112007 (2016)

Predicted number of signal events, given (an older version of) Geant4 simulation constrained with external data:

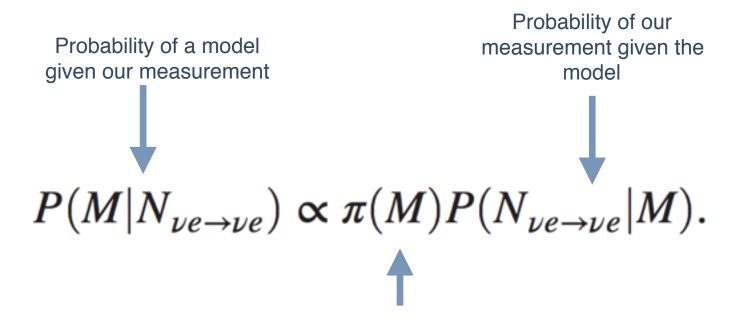
149 ± 19

Observed in Data: 
$$135 \pm 17.0$$

How to combine this with our existing flux model?



### We use a Bayesian argument:



Prior probability of the model



We use a Bayesian argument:

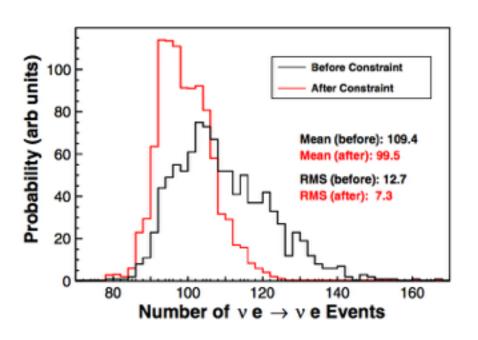
$$P(M|N_{\nu e \to \nu e}) \propto \pi(M)P(N_{\nu e \to \nu e}|M).$$

We estimate this by computing a chi-square between the measured electron energy distribution and the prediction for the universe in question

$$P(N_{\nu e \to \nu e} | M) = \frac{1}{(2\pi)^{K/2}} \frac{1}{|\Sigma_{\mathbf{N}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{N} - \mathbf{M})^T \Sigma_{\mathbf{N}}^{-1}(\mathbf{N} - \mathbf{M})}$$



An example using the total number of predicted neutrino-electron scattering events:



Each entry in the "before constraint" distribution corresponds to a "flux universe"

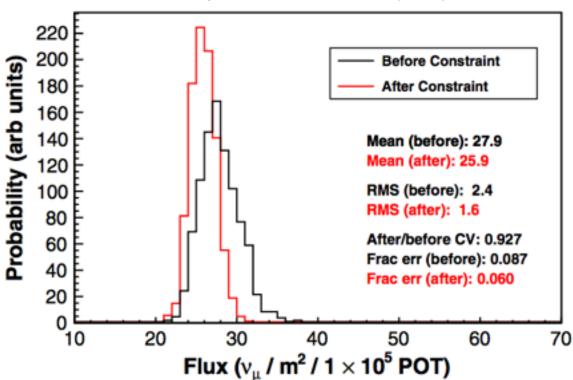
To produce "after constraint", each original entry is weighted by the probability of that universe given the neutrino-electron scattering measurement

$$P(N_{\nu e \to \nu e} | M) = \frac{1}{(2\pi)^{K/2}} \frac{1}{|\Sigma_{\mathbf{N}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{N} - \mathbf{M})^T \Sigma_{\mathbf{N}}^{-1}(\mathbf{N} - \mathbf{M})}$$



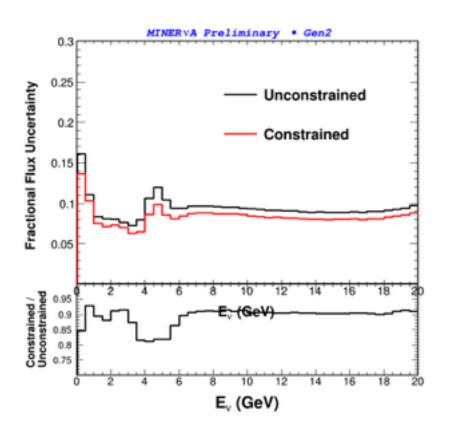
The same weights can be applied to constrain the flux uncertainty any quantity predicted by our simulation:

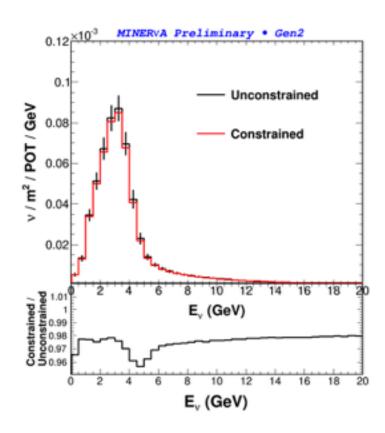




Probability
distributions for total
muon neutrino flux
integrated between
0-10 GeV, before and
after constraint



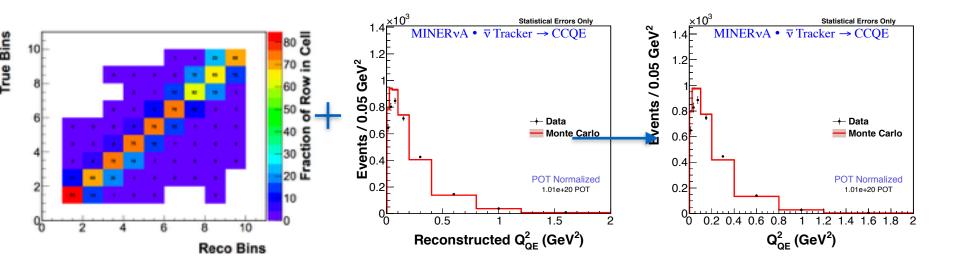




 This statistically limited result reduces MINERvA's flux uncertainties as a function of energy by 10-20% (of the a priori uncertainties)



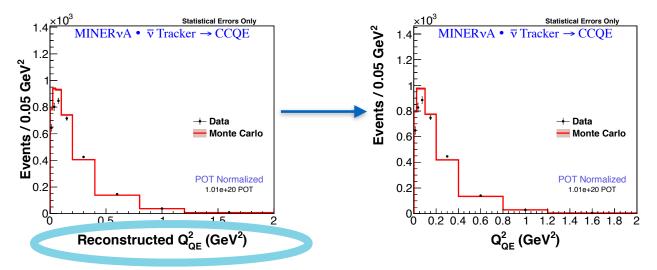
• I promised you I would say more about unfolding:



$$\left(\frac{d\sigma}{dQ_{\text{QE}}^{2}}\right)_{i} = \frac{\sum_{j} \left(M_{ij} \left(N_{\text{data},j} - N_{\text{data},j}^{\text{bkgd}}\right)\right)}{\epsilon_{i} \left(\Phi T\right) \Delta Q_{\text{QE},i}^{2}}$$



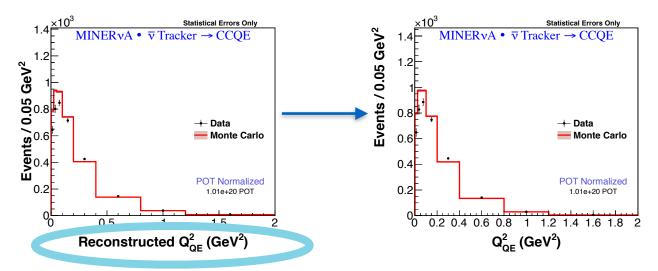
 We try hard to define analyses such that we are unfolding to correct for detector smearing and not for highly model-dependent effects such as final state interactions:



In principle, Q<sup>2</sup> is the 4-momentum transferred from the neutrino to the final state nucleon



 We try hard to define analyses such that we are unfolding to correct for detector smearing and not for highly model-dependent effects such as final state interactions:



In practice (for this analysis), we approximate Q<sup>2</sup> using measurements of the final state muon's energy and angle

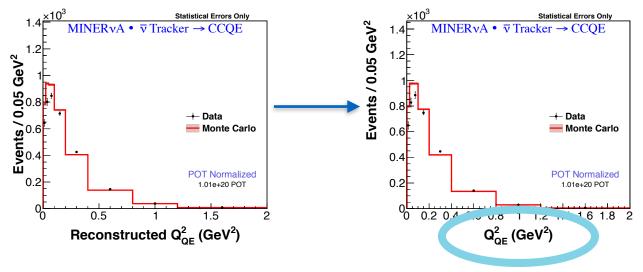
$$E_{\nu}^{QE} = \frac{m_n^2 - (m_p - E_b)^2 - m_{\mu}^2 + 2(m_p - E_b)E_{\mu}}{2(m_p - E_b - E_{\mu} + p_{\mu}\cos\theta_{\mu})}$$

$$Q_{QE}^2 = 2E_{\nu}^{QE}(E_{\mu} - p_{\mu}\cos\theta_{\mu}) - m_{\mu}^2,$$

Even if we could perfectly reconstruct the muon variables, this differs from the original interaction Q<sup>2</sup> due to initial state nucleus effects



 We try hard to define analyses such that we are unfolding to correct for detector smearing and not for highly model-dependent effects such as final state interactions:

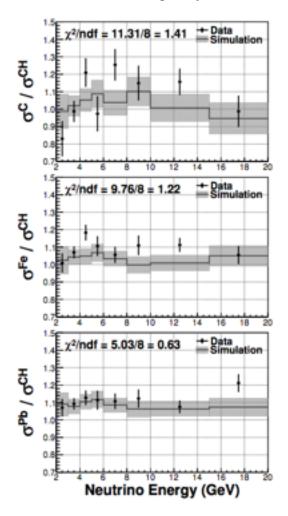


So we don't unfold to the original Q<sup>2</sup> of the nucleon interaction, but the same quantity we reconstruct, with reconstructed muon momenta replaced by true muon momentum leaving the nucleus

$$\begin{split} E_{\nu}^{QE} \; &= \; \frac{m_n^2 - (m_p - E_b)^2 - m_{\mu}^2 + 2(m_p - E_b)E_{\mu}}{2(m_p - E_b - E_{\mu} + p_{\mu}\cos\theta_{\mu})} \\ Q_{QE}^2 \; &= \; 2E_{\nu}^{QE}(E_{\mu} - p_{\mu}\cos\theta_{\mu}) - m_{\mu}^2, \end{split}$$



This is not always possible:



For our charged current inclusive measurements, we reconstruct neutrino energy by adding muon energy to calorimetricaly corrected recoil

We have to rely heavily on models to tell us the relationship between the recoil energy we see in the detector and the true recoil energy (e.g. the number of neutrons that leave the detector)



We use iterative bayesian unfolding as implemented in RooUnfold

#### **Bayesian Method**

The Bayes' theorem is used repeatedly to get the best estimates of the true distribution [1]. Considering:

$$\hat{n}(C_i) = \sum_{j=1}^{n_E} M_{ij} n(E_j)$$

 \hat{n}(C\_i): Unfolded distribution.

M<sub>\*</sub> : Unfolding Matrix

n(E): Folded distribution

The Bayesian method is used for calculating the unfolding matrix:

$$M_{ij} = \frac{P(E_j|C_i)n_0(C_i)}{\epsilon_i f_j}$$

P(E,|C|): Migration Matrix

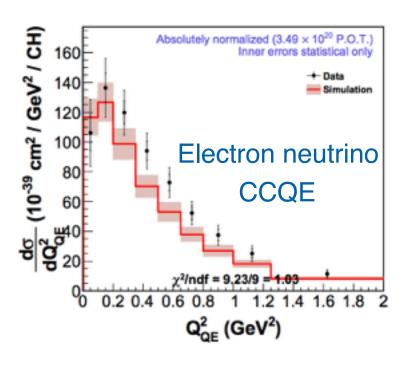
f : Efficiencies : folded prior distribution

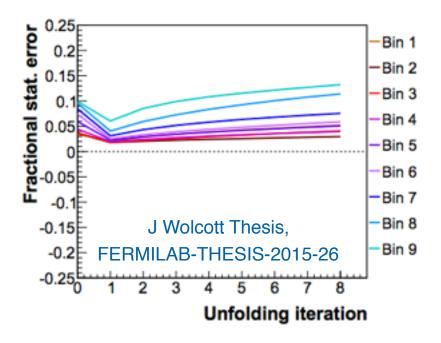
 $n_0(C_i)$ : arbitrary, then updated from previous  $\hat{\pi}(C_i)$  it.

[1]A multidimensional unfolding method based on Bayes' theorem, G.D' Agostini, NIM-A Vol. 362, No. 2-3. (15 August 1995)



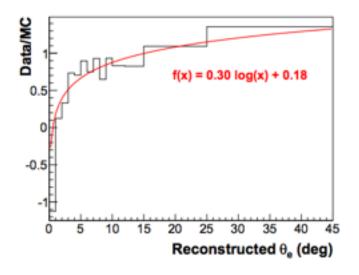
- The big question for every analysis: how many iterations?
  - More iterations give you less model dependence but higher statistical uncertainties

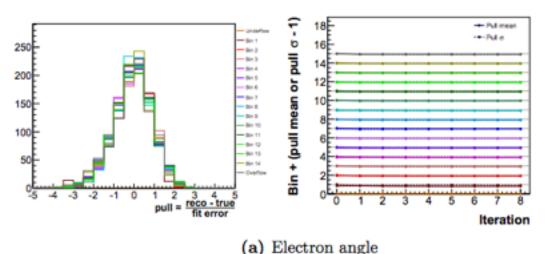






- The big question for every analysis: how many iterations?
  - Each analyzer does a study where they warp the underlying MC distribution and study how many iterations are required to 'recover' the original MC distribution

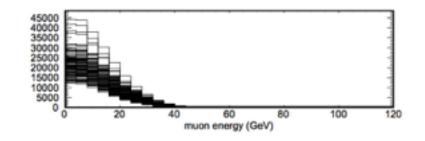




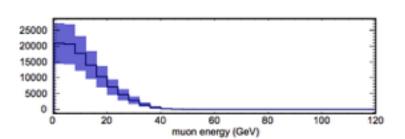
# **Unfolding Systematic Uncertainties**

 Systematic uncertainties due to model dependence of the unfolding is assessed by performing unfolding in all of the varied systematic universes

In principle, you'd want to vary the unfolding matrix for all systematics variation

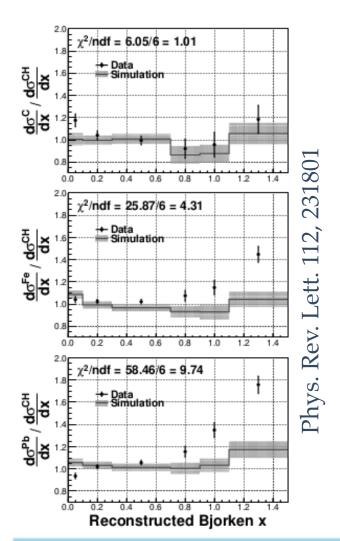


In practice, we find this inflates systematic uncertainties with statistical fluctuations; we generally only vary the unfolding matrix in cases where we expect the variation to impact the matrix





And in some cases we don't unsmear at all:

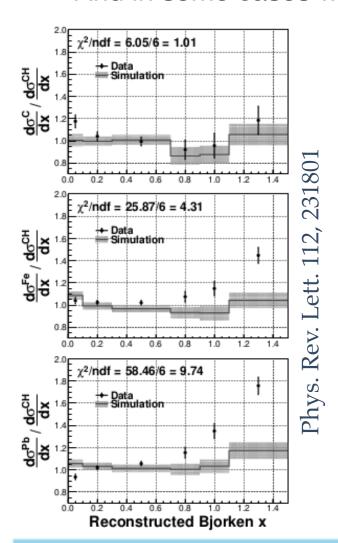


For example, charge current inclusive ratios across nuclear targets as a function of x, which has large amounts of smearing

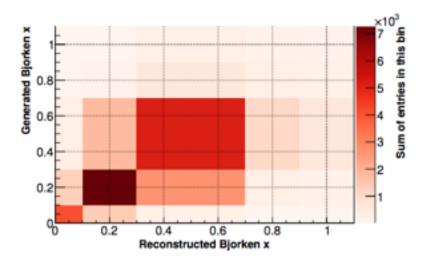
$$x=rac{Q^2}{2M
u}$$
 high x = more elastic  $u=E_
u-E_\mu$   $Q^2=2E_
u\left(E_\mu-p_\mu cos\left( heta_\mu
ight)
ight)$ 



And in some cases we don't unsmear at all:



For example, charge current inclusive ratios across nuclear targets as a function of x, which has large amounts of smearing



(b)  $x_{bj}$ , Lead of Target 4

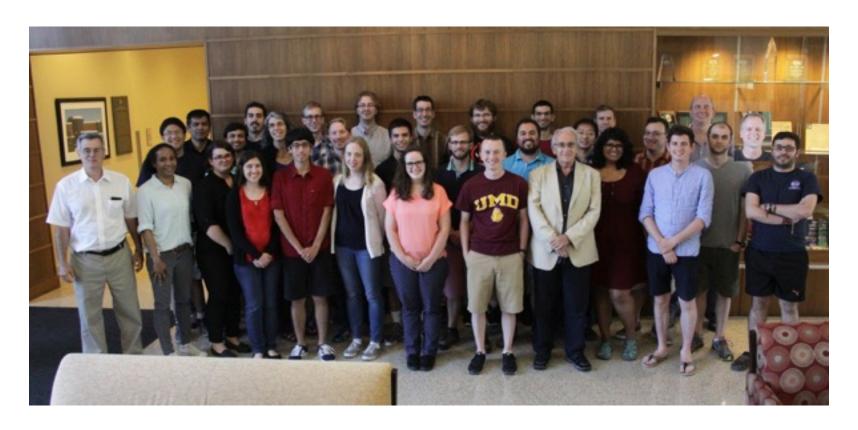


### Conclusion

- MINERvA has lots of data
- With it comes a lot of statistical challenges
- We try to do a good job of meeting them, in spite of none of us being statisticians
- In some cases (e.g. flux constraint), we are developing techniques that are likely to be useful to future oscillation experiments
- Your comments are welcome!



# From the MINERvA Collaboration:



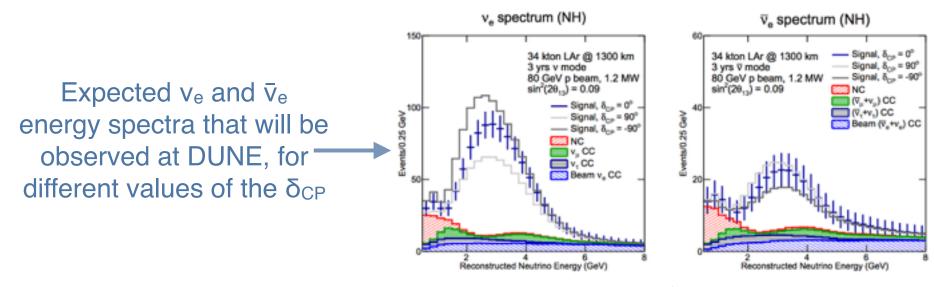
# Thank You!!



# **Searching for Rare Processes: Coherent Kaon Production**

### Introduction: Why

MINERvA makes measurements that support long-baseline experiments such as DUNE

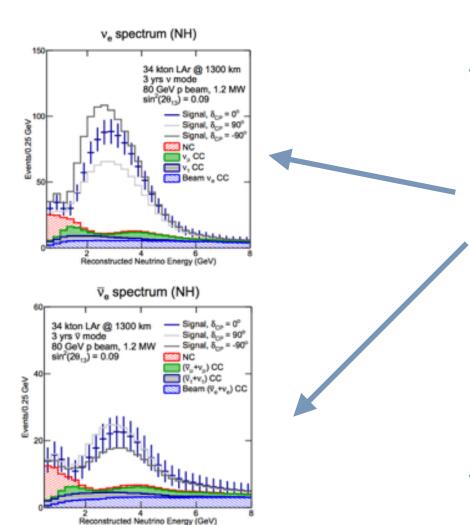


LBNE arXiv:1307.7335

 DUNE will make many of its measurement by comparing what they see with predictions



### **Introduction: Why**

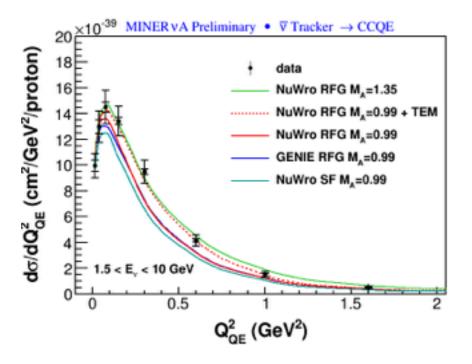


- To produce these predictions, we need a detailed model of neutrino interactions with matter
  - \* list of all the types of neutrino interaction processes that can occur in the detector
  - The probability that each process will happen (which we call the crosssection)
  - What they look like when they do
- This is one of the biggest source of systematic uncertainty for experiments like DUNE

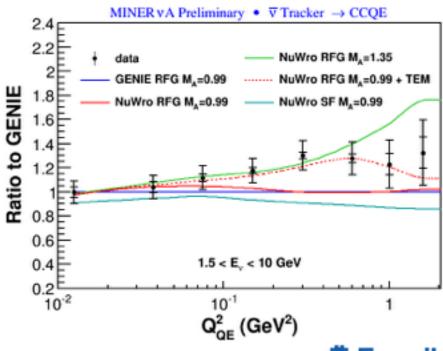


### **Introduction: How**

 We typically compare our final cross sections to models, and make the data available to future model tuners:



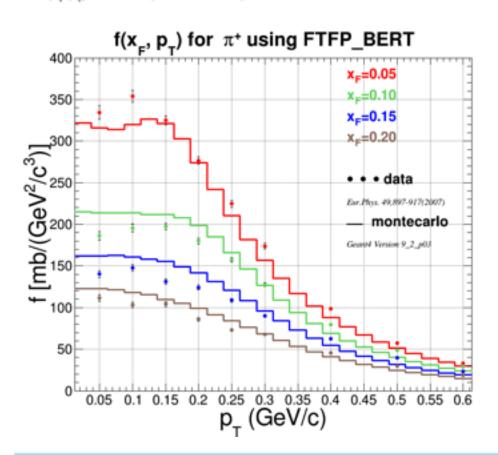
Phys. Rev. Lett. 111, 022501 (2013) arXiv:1305.2234



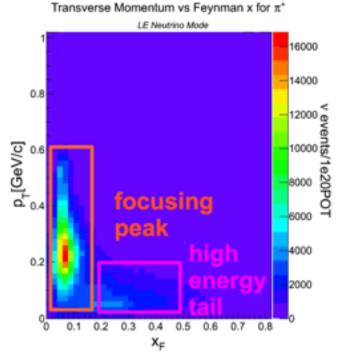
# **Constraining the NuMI Flux**

One problem: Geant4 does not always agree with external data:

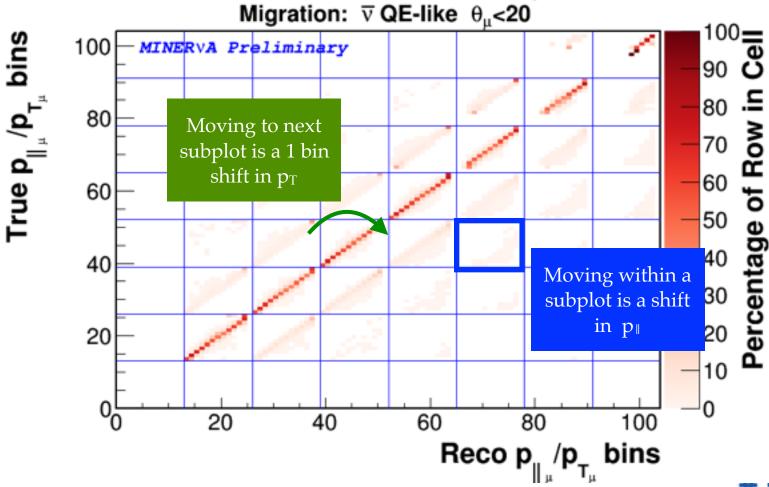
$$f(x_r, p_r) = E d^3\sigma/dp^3 = invariant production cross-section$$



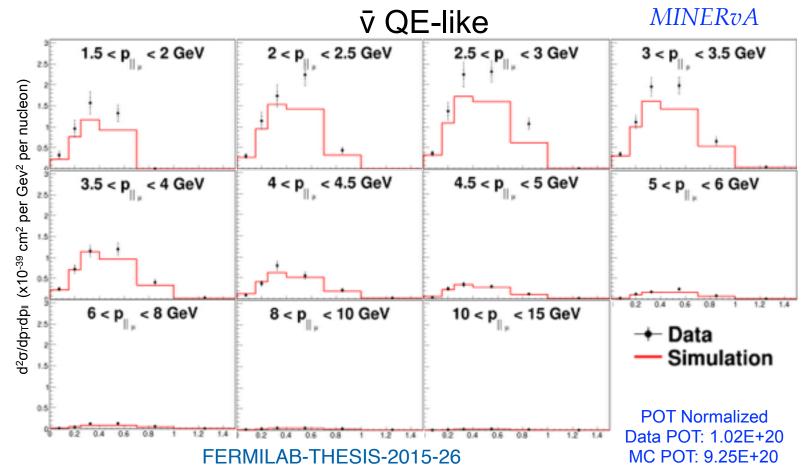
$$x_F = 2\frac{P_L}{E_{cm}}$$



 Unfolding becomes and even greater challenge for analyses measuring two dimensional cross sections:

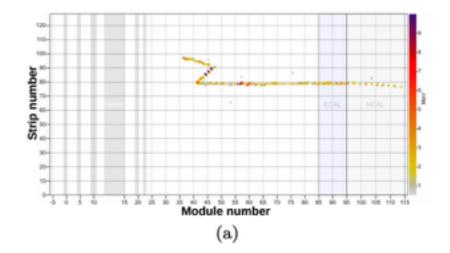


 Unfolding becomes and even greater challenge for analyses measuring two dimensional cross sections:



### **Kaon Production**

- Kaon production by neutrinos is interesting because it is a background to proton decay measurements
- One potential source of kaon production is coherent kaon production
  - Veeery small cross section never seen before
  - But MINERvA went looking for it



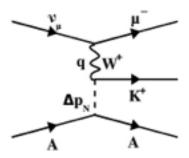
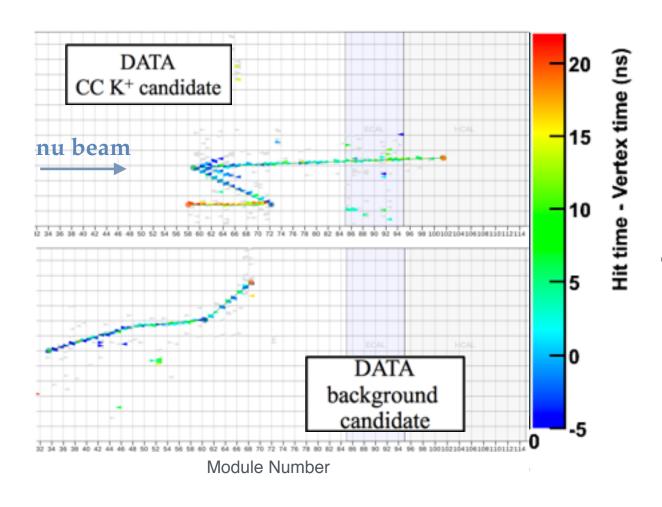


FIG. 1: Feynman diagram for coherent charged kaon production. The square of the momentum transfer to the nucleus is  $|\Delta p_N|^2 = |q - p_K|^2 = |t|$ .

Phys. Rev. Lett. 117, 061802 (2016)



### **Kaon Production**



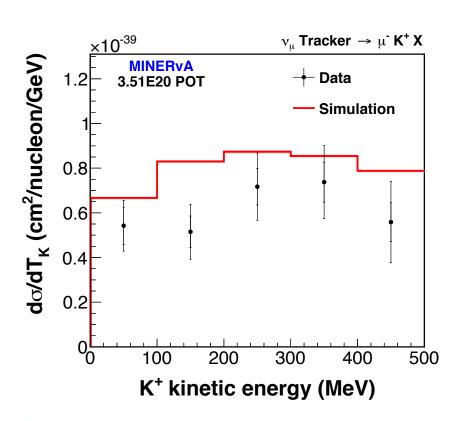
Key distinguishing feature of kaons for MINERvA: time separation of kaon and decay products

Here, color denotes hit time

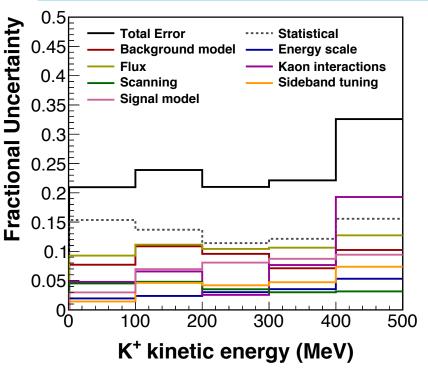


### **Kaon Production**

- Charged current K+ production cross section shows reasonably good agreement with simulation.
- This measurement increased the world's sample of K+ production events from neutrinos from dozens to thousands!



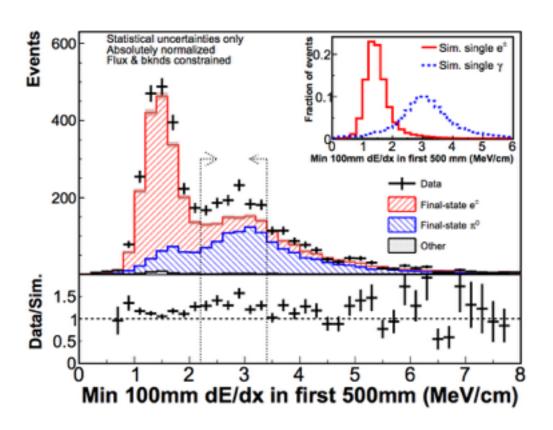
### https://arxiv.org/abs/1604.03920





### **Neutral Current Diffractive Pion Production**

arXiv:1604.01728, Phys. Rev. Lett. 116, 081802 (2016)



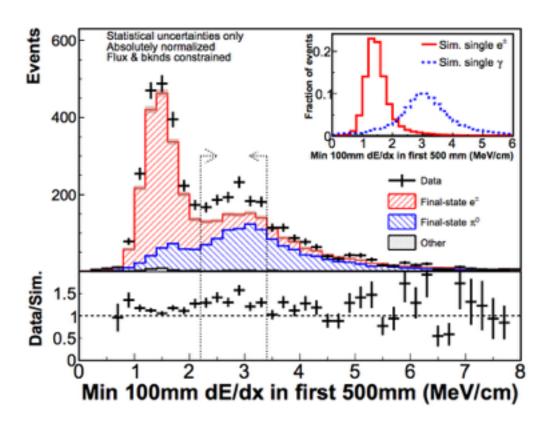


Sometimes when we go hunting for the golfballs of oscillation experiments...



### **Neutral Current Diffractive Pion Production**

arXiv:1604.01728, Phys. Rev. Lett. 116, 081802 (2016)

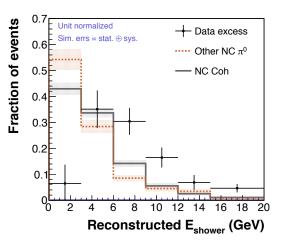


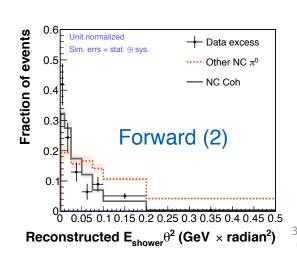


... we also find alligators!

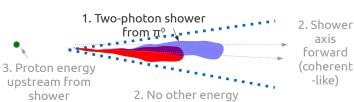


### **Neutral Current Diffractive Pion Production**

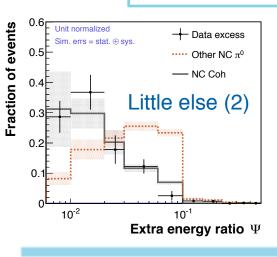


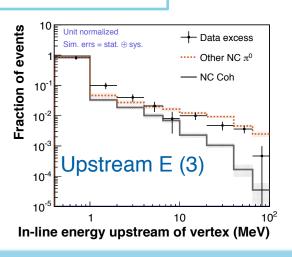


- Two-photon Shower
- Coherent-like scattering
  - **Forward Kinematics**
  - Very little other energy
- 3) Visible proton energy









NC diffractive  $\pi^0$  production from Hydrogen



